昆 虫 学 报 *Acta Entomologica Sinica*, September 2013, 56(9): 1063 – 1074 ISSN 0454-6296

Cloning and bioinformatical analysis of vitellogenin gene of the Indian malaria vector *Anopheles culicifacies* (Diptera: Culicidae)

Monika MIGLANI, Surendra Kumar GAKHAR*

(Centre for Biotechnology, Maharshi Dayanand University, Rohtak, Haryana 124001, India)

Abstract: Vitellogenin (Vg) is the major yolk protein precursor which is synthesized abundantly in the insect fat body after the female ingests blood meal. The regulatory elements of vitellogenin have been used to drive the tissue specific expression of anti parasitic gene in mosquitoes, where its maximum interaction could take place with the parasite. However, no endogenous promoter has been analysed so far in the Indian malaria vector *Anopheles culicifacies* which is responsible for 60% – 70% of malaria cases in India. In this study, the vitellogenin gene including 5′ upstream regulatory region of *Anopheles culicifacies* was cloned after PCR amplification and named *AncuVg* (GenBank accession number JN113091). It contains an ORF of approximately 6.2 kb encoding 2 052 amino acids with a putative signal peptide of 16 residues. It also contains an N_Vitellogenin region and a VWF type D domain, that are found conserved in other insect Vgs too. The molecular weight of the predicted polypeptide is 238.0 kDa. It possesses four consensus (RXXR/S) cleavage sites and close to the C-terminus there is a GL/ICG motif followed by nine cysteine residues and a DGXR motif, located 18 residues upstream from the GL/ICCG motif. Three polyserine regions were found in the deduced amino acid sequence: two in the amino terminal region and one in the carboxy terminal region. The extent of codon bias in mosquito vitellogenin genes based on the relative synonymous codon usage values were determined by the effective number of codons. The 3D structure of *A. culicifacies* Vg was also predicted. The 5′ upstream region of the *AncuVg* gene was analyzed to understand the regulation of Vg gene transcription. Phylogenetic analysis using the 5′ upstream region of Vg genes showed their conformation to three major clades among mosquitoes. Homology and other characteristic features of Vg have also been analyzed using various bioinformatic tools.

Key words: *Anopheles culicifacies*; vitellogenin; mosquito; phylogenetic analysis; polyserine; transcription factor

1 INTRODUCTION

Malaria alone accounts for more than 200 million acute illnesses per year with near to one million deaths annually (World Health Organisation, 2011). The development of drug resistance in *Plasmodium* and insecticide resistance in mosquitoes coupled with the lack of efficient vaccines necessitate the development of novel strategies to combat malaria.

Anopheles culicifacies is one of the major vectors of malaria in India accounting for 60 – 70% of malaria cases in the country. It exists as a species complex comprised of five sibling species which have been provisionally denominated as species A, B, C, D and E (Green and Miles 1980; Subbarao *et al.*, 1983; Suguna *et al.*, 1989; Vasantha *et al.*, 1991; Kar *et al.*, 1999). The advent of mosquito germ line transformation made the creation of transgenic mosquitoes impaired in malaria transmission a reality (Ito *et al.*, 2002; Moreira *et al.*, 2002). An indispensable step in engineering such mosquitoes with diminished vector competence is the recognition of appropriate promoters to carry on the expression of the effector genes (Riehle *et al.*, 2003; Nirmala *et al.*, 2005). Endogenous mosquito promoter and other *cis*-acting DNA sequences are required to direct the optimal tissue-, stage- and sex-specific expression of the effector molecules (Nirmala *et al.*, 2005).

Vitellogenin (Vg), the major yolk protein precursor, is synthesized in the fat body (analogous to the vertebrate liver) and highly expressed after the female ingests blood meal (Ahmed *et al.*, 2001; Kokoza *et al.*, 2001). It is secreted into the haemolymph, and subsequently sequestered in the developing oocytes through receptor-mediated endocytosis (Byrne *et al.*, 1989; Raikhel and

* Corresponding author, Tel.: +919896014951; E-mail: surengak@gmail.com

Received: 2013-05-06; Accepted: 2013-07-18

Dhadialla, 1992; Sappington and Raikhel, 1995). Mosquito Vg is highly expressed after a blood meal, therefore Vg control sequences should be the most useful for targeting an effector molecule to the sporozoites in the haemolymph (Ito *et al.*, 2002). In spite of the potential of Vg *cis*-regulatory elements for the development of mosquito transgenesis-related technologies, no endogenous promoter has so far been reported in *Anopheles culicifacies*.

The Vgs have been sequenced from 29 insect species so far belonging to six distinct orders (Tufail and Takeda, 2008; Chen *et al.*, 2010; Li *et al.*, 2010). Earlier investigators have analyzed the phylogenetic relationship of vitellogenins among various insect species (Nose *et al.*, 1997; Lee *et al.*, 2000b; Tufail *et al.*, 2007; Chen *et al.*, 2010), gene duplication and amino acid composition in mosquitoes (Isoe and Hagedorn, 2007). Most of these studies are based on either nucleotide sequences or amino acid sequences. No study appears to have been made on phylogenetic analysis based on 5′ upstream regulatory sequences. However, 5′ upstream regulatory sequences of only two mosquitoes, *i. e.*, *A. stephensi* and *A. gambiae* have been used for transgenesis (Nirmala *et al.*, 2006; Chen *et al.*, 2007). Therefore, in the present study the vitellogenin gene of *A. culicifacies* (*AncuVg*) has been isolated, sequenced and characterized including its upstream regulatory region to understand the regulation of Vg gene transcription. The phylogenetic analysis was also carried out using 5′ upstream regulatory sequences. The regulation of expression of highly expressed proteins sometimes involves the use of specific synonymous codons and also studies of synonymous codon usage can reflect information about the molecular evolution of an individual gene. In addition, the synonymous codon usage of Vg genes has also been analyzed that resulted into significant codon usage bias.

2 MATERIALS AND METHODS

2.1 Mosquito rearing

The cyclic colonies of *A. culicifacies* Species A (Dhera Strain) mosquitoes were maintained in an insectary at temperature of 28 ± 2℃ and 70% – 80% relative humidity with a photoperiod of 14 hour light and 10 hour dark (Gakhar *et al.*, 1997). Dawn and Dusk effect was necessary for stimulating them for mating. Adult mosquitoes kept in 30 cubic cm cloth cages were fed on 1% glucose solution and water soaked raisins. Females were fed 3 – 4 days after eclosion on anesthetized mice when a blood meal was required for ovarian development. On the 3rd day post blood feeding females were allowed to lay eggs in water filled plastic bowls lined with filter paper. Larvae were reared in enamel trays at a standard density of 300 larvae/450 mL of water and were fed on yeast extract and dog biscuits in the ratio of 2∶3 (w/w). It passes through 4 instars for about 12 – 13 days. After pupation, the pupae were transferred to fresh bowl and were kept in cloth cages for emergence to adult mosquitoes (Gakhar *et al.*, 2001).

2.2 Genomic DNA extraction

Genomic DNA was isolated from *A. culicifacies* female mosquito by using standard phenol-chloroform method (Sambrook *et al.*, 1989). Typically one individual mosquito was homogenized in Bender buffer (0.1 mol/L NaCl, 0.2 mol/L scrose, 0.1 mol/L Tris (pH 9.0), 0.05 mol/L EDTA (pH 8.0), 0.5 mol/L SDS). RNA contamination was removed by the addition of RNAse at 37℃ to a final concentration of 0.2 mg/mL. The samples were then treated with Proteinase K to a final concentration of 0.6 mg/mL to digest all protein matter and incubated at 50℃ overnight.

2.3 Cloning and sequencing

Oligonucleotide primers were designed based on the highly conserved regions of already published genomic sequence of *A. gambiae* Vg and *A. stephensi* Vg (GenBank accession numbers: AF281078 and DQ442990, respectively) using Primer3Plus (Table 1). The standard PCR reaction was carried out for 5 min at 95℃, followed by 35 cycles consisting of 95℃ for 1 min, 55℃ for 45 s, 72℃ for 1 min and then held at 72℃ for 10 min. PCR products were separated on a 1% agarose gel in 1 × TAE buffer. PCR fragments were purified using PCR Purification Kit (Genei) and cloned into the TOPO TA cloning vector (Invitrogen). Clones were sequenced using M13-forward and M13-reverse universal primers, in an automated fluorescence sequencing system 3730 DNA Analyzer (Applied Biosystems).

2.4 Analysis of amino acid sequence of *AncuVg*

Analysis of signal peptides in the amino acid sequence of *AncuVg* was performed with the SignalP 4.0 program (http://www.cbs.dtu.dk/services/SignalP/). The prediction programs for post-translational modifications of glycosylation and phosphorylation were NetNGlyc 1.0 and NetPhos 2.0, respectively. The isoelectric point (pI) and molecular weight (MW) were computed by Compute pI/MW (http://www.expasy.ch/tools/pi_tool.html). Amino acid composition analysis was performed

Table 1　Primers used in this study

No.	Primer name	Primer sequence (5′-3′)
1.	VGF1	CTGTAAACATGTGGGCAGTAAAA
	VGR1	ACGGTCAATGTAGGCAACGA
2.	VGF2	AACGTGACGACGAAAACCAT
	VGR2	AGACGTCACCAGCAGGAAGTA
3.	VGF3	CAACCGTCGTGATCTGAATG
	VGR3	CTTGTTTCCGAACTGCTCGT
4.	VGF4	AGCGTCTGGCTTTTCATTGTC
	VGR4	AAGTCGAAGCCAAAGTCGAA
5.	VGF5	TTCGACTTTGGCTTCGACTT
	VGR5	CATCAACAGACGCTCCTTCTC
6.	VGF6	AGCGTCTGTTGATGTTCCTG
	VGR6	GACCGACTTGTTGCTCATGT
7.	VGF7	CAACAAGTCGGTCACCAAGA
	VGR7	GGGGGCATCAAGAAAAGATG

using MEGA 5.0 (Tamura *et al.*, 2011). The 3D model was obtained by submitting the AncuVg protein sequence to the ESyPred3D server (Lambert *et al.*, 2002). The template used by ESyPred3D was the structure having the following PDB identifier: 1LSH chain 'A', the structure of a lipoprotein (lipovitellin) from *Ichtyomyzon unicuspis*.

2.5　Codon usage indices analysis

Relative synonymous codon usage (RSCU) values of each codon of the gene were used to measure the synonymous codon usage (Sharp *et al.*, 1986). RSCU values are precisely useful in comparing codon usage between genes, or sets of genes that differ in their size and amino acid composition and are mostly independent of amino acid composition. The preferred codon usage of each gene was analyzed using CAIcal software (Pubaigo *et al.*, 2008). The effective number of codons (ENC) was used to quantify the codon usage bias of each gene; which is the best overall estimator of absolute synonymous codon usage bias (Wright, 1990). The GC index was used to calculate the overall GC content in vitellogenin gene of each organism, while the index GC_{3s} was used to calculate the fraction of GC nucleotides at the synonymous third codon position (Richard *et al.*, 2000).

2.6　Analysis of binding motif at the 5′ upstream of *AncuVg* gene

The 5′ upstream sequence of the *AncuVg* gene was analyzed to understand the regulation of Vg gene transcription. Consensus sequences of the response elements were determined by a manual search as follows: (A/T) GATA (A/G) for GATA factor (Orkin, 1992). Consensus sequences of ecdysone response elements (EcRE) were represented by (A/G)G(G/T)T(C/A)A(N)TG(C/A)(C/A)(C/t)(C/T); for other ecdysone early genes were: (C/A)GGAA for E74A and B (Urness and Thummel, 1995); broad complex (BRC) isoform specific response elements were as follows: TA(T/A)(T/A)(A/G)ACAA(A/G)(T/A) (Z1), TT(T/A)(T/A)CTATTT (Z2), (T/A)AAAC(T/A)(T/A)(A/G)(T/A) (Z3) and (A/G)(T/G)AAA(C/G)A (Z4) (von Kalm *et al.*, 1994). Motif search of the upstream region of the Vg gene was also analyzed using CONSITE software (Sandelin *et al.*, 2004).

2.7　Sequence comparisons and phylogenetic analysis

Vitellogenin nucleotide sequences used for comparisons were obtained from GenBank database. GenBank accession numbers of the sequences used are listed below (Table 2). Sequence alignments were performed using nBLAST http://www.ncbi.nlm.nih.gov/blast/. Homology of *AncuVg* was compared with the nucleotide sequences of Vg genes from different mosquitoes. Multiple alignments were performed by the CLUSTAL W program (Thompson *et al.*, 1994). Phylogenetic tree of the 5′ upstream region of mosquito Vg genes was profiled by the Neighbor-Joining method (Saitou and Nei, 1987) with bootstrap values assessed at 1 000 replicates (Felsenstein, 1985). The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*, 2004) and were in the units of the number of base substitutions per site. The analysis involved 21 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 103 positions in the final data set. Evolutionary analyses were conducted in MEGA 5.0.

3　RESULTS

3.1　Isolation and structure of the *AncuVg* gene

The sequence analysis of vitellogenin gene of *A. culicifacies* (deposited in GenBank under the accession number JN113091) revealed the presence of 3 exons and 2 introns (Fig. 1). The total combined length of the 3 exons was 6 159 bp including relatively long second exon of 4 929 bp in length. The *AncuVg* gene contains an ORF of approximately 6.2 kb. The ORF has both a start (ATG) and stop codon (TAA), indicating that the sequence contains the complete coding region. A

Table 2 GenBank accession numbers of the sequences used and percent identity of *AncuVg* gene sequence with other mosquito Vg genes

No.	Organism	Gene	GenBank accession no.	Identity to *AncuVg* (%)
1	*Anopheles gambiae*	*Vg1*, *Vg2*	AF281078	91
2	*Anopheles stephensi*	*Vg1*, *Vg2*	DQ442990	93
3	*Anopheles albimanus*	*VgC*	AY691327	87
4	*Aedes aegypti*	*VgA1*	L41842	69
5	*A. aegypti*	*VgB*	AY380797	69
6	*A. aegypti*	*VgC*	AY373377	73
7	*Aedes albopictus*	*VgA1*	AY691316	No significant identity
8	*A. albopictus*	*VgC*	AY691317	70
9	*Aedes polynesiensis*	*VgA1*	AY691318	No significant identity
10	*A. polynesiensis*	*VgB*	AY691319	69
11	*A. polynesiensis*	*VgC*	AY691320	70
12	*Culex tarsalis*	*Vg1a*	GU017909	76
13	*C. tarsalis*	*Vg1b*	GU017910	76
14	*C. tarsalis*	*Vg2a*	GU017911	66
15	*C. tarsalis*	*Vg2b*	GU017912	67
16	*Culex quinquefasciatus*	*VgC1*	AY691324	75
17	*C. quinquefasciatus*	*VgC2*	AY691325	70
18	*Ochlerotatus atropalpus*	*VgB*	AY691321	67
19	*O. atropalpus*	*VgC*	AY691322	69
20	*Ochlerotatus triseriatus*	*VgC*	AY691323	No significant identity
21	*Toxorhynchites amboinensis*	*VgC*	AY691326	69

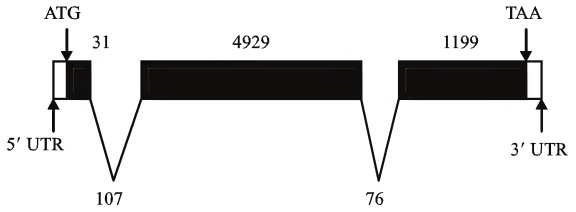

Fig. 1 Structure diagram of the Vg gene from *Anopheles culicifacies* (*AncuVg*)
Exons are indicated by black and white gaps represents introns. UTRs are indicated by arrow.

putative polyadenylation signal is located at nucleotide positions 6 757 – 6 769.

3.2 Intron analysis

The introns found in mosquito vitellogenin genes varied in length. The 1st intron ranged from 62 bp (*Cx. tarsalis*) to 119 bp (*A. stephensi*) and the 2nd intron ranged from 57 (*Ae. aegypti*) to 84 bp (*A. stephensi*). The splice sites for the two introns of *AncuVg* was invariably characterized as the 1st intron in particular had G/GTAAGT and TTACAG/T at the 3′ and 5′ splice sites, respectively, while the 2nd intron had C/GTAAGT and TTTCAG/C at the 3′ and 5′ splice sites, respectively (Table 3).

3.3 Characterization of the deduced amino acid sequence

The conceptual translation of the nucleotide sequence indicated that ORF encoded 2 052 amino acids that contained a Vitellogenin _ N region (lipoprotein amino terminal region) and Von Willebrand factor (VWF) type D domain. The first 16 amino acids (aa) corresponded to a signal peptide as predicted by the SignalP (Nielsen *et al.*, 1997). The predicted molecular weight of *AncuVg* after cleavage of the signal peptide was 238.0 kDa. There were four putative cleavage sites showing the RXXR consensus sequence and thus the putative protein could potentially be cleaved into nine subunits (Fig. 2).

Twenty putative glycosylation sites were predicted, however the sites placed at positions 32, 50, 285, 316, 442 and 1 071 may be effectively glycosylated according to the NetNGlyc 1.0 Prediction Server. In addition, 178 phosphorylation sites were also located by NetPhos 2.0 Prediction Server. It also included three poly-serine region's characteristics of vertebrate and several insect Vgs: two in the amino-terminal region (365 – 386 aa and 445 – 463 aa) and one in the carboxy-terminal ends (1 917 – 1 944 aa).

The conserved motif GL/ICG at position 1 829 – 1 832, is followed by a number of cysteines (nine in

Table 3　Exon lengths and intron boundaries of complete mosquito vitellogenin genes

Name of organism	Gene	Exon 1 (bp)	Intron 1		Exon 2 (bp)	Intron 2		Exon 3 (bp)
			3′ Splice sites	5′ Splice sites		3′ Splice sites	5′ Splice sites	
Anopheles culicifacies	*Vg1*	31	G/GTAAGT ··· (107) * ···	TTACAG/T	4 929	C/GTAAGT ··· (76) ···	TTTCAG/C	1 199
Anopheles gambiae	*Vg1*	31	G/GTAAGT ··· (99) ···	TTACAG/T	4 941	C/GTAAGT ··· (74) ···	CTACAG/C	1 184
Anopheles stephensi	*Vg1*	31	G/GTAAGT ··· (119) ···	TTGCAG/T	4 947	C/GTAAGT ··· (84) ···	CTCCAG/C	1 199
Anopheles albimanus	*VgC*	31	G/GTAAGT ··· (101) ···	CTGCAG/T	4 881	C/GTAAGT ··· (83) ···	CATCAG/C	1 178
Aedes aegypti	*VgA1*	31	G/GTAAGT ··· (70) ···.	CCACAG/C	5 214	C/GTAAGT ··· (57) ···	TTTCAG/C	1 202
A. aegypti	*VgB*	31	G/GTAAGT ··· (82) ···.	CCACAG/T	5 199	C/GTAAGT ··· (59) ···	TTGCAG/C	1 199
A. aegypti	*VgC*	31	G/GTAAGT ··· (67) ···.	CAATAG/T	4 995	C/GTAAGT ··· (68) ···	TCGCAG/C	1 238
Ochlerotatus atropalpus	*VgB*	31	G/GTAAGT ··· (74) ···.	TCGCAG/T	5 190	C/GTAAGT ··· (61) ···	TTCTAG/C	1 103
O. atropalpus	*VgC*	31	G/GTCAGT ··· (80) ···.	TTTCAG/T	4 992	C/GTAAGT ··· (68) ···	TCTTAG/C	>453
Culex tarsalis	*Vg1a*	31	G/GTAAGT ··· (63) ···.	ACGCAG/T	5 061	T/GTAAGA ··· (61) ···	CTACAG/C	1 262
C. tarsalis	*Vg1b*	31	G/GTAAGT ··· (64) ···.	ACGCAG/T	5 067	T/GTAAGA ··· (59) ···	CTACAG/C	1 262
C. tarsalis	*Vg2a*	31	G/GTAAGC ··· (62) ···.	TTCCAG/T	5 013	C/GTAAGT ··· (75) ···	TTCCAG/C	1 199
C. tarsalis	*Vg2b*	31	G/GTAAGC ··· (62) ···.	TTCCAG/T	5 028	C/GTAAGT ··· (75) ···	TTCCAG/C	1 199

* Size of the intron is mentioned in parentheses.

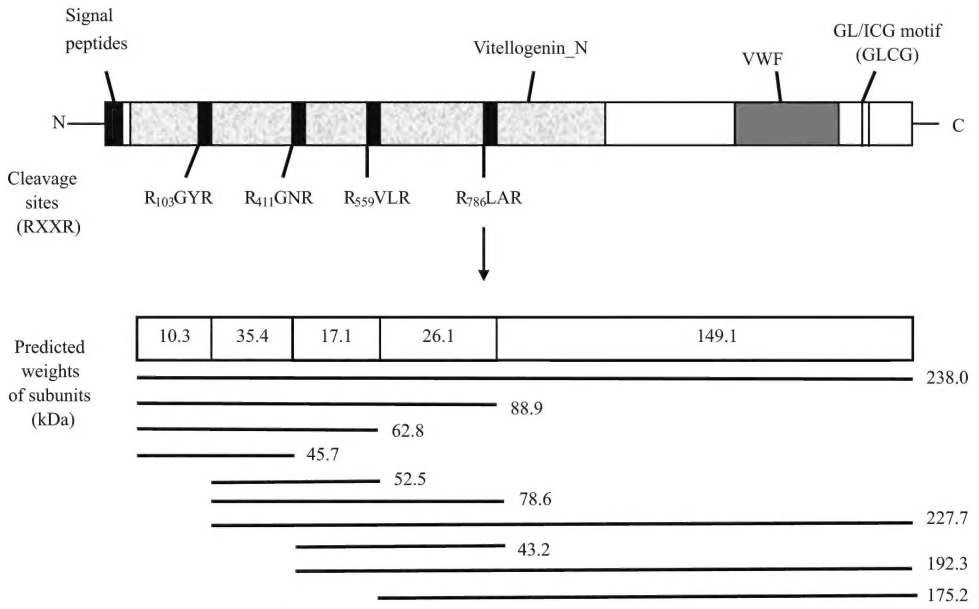

Fig. 2　Predicted structure and molecular weights (kDa) of *Anopheles culicifacies* vitellogenin (AncuVg) peptide

AncuVg is speculated to be cleaved by RXXR sites at amino acid residues 103, 411, 559 and 786. The numbers indicate molecular weights of native and cleaved subunits of the gene. VWF: von Willebrand factor.

hymenopterans) at conserved locations near the C-terminus (positions 1 848, 1 867, 1 886, 1 965, 1 981, 1 990, 1 994, 2 008 and 2 048). In addition to this, the DGXR motif is located 18 residues upstream of the GL/ICG motif (Fig. 3). Furthermore, there are no significant differences in the amino acid composition of mosquito vitellogenin genes (data not shown). Most are rich in tyrosine (Y) and phenylalanine (F); however, Vg2a and Vg2b of *Cx. tarsalis* and VgC1 and VgC2 of *Cx. quinquefasciatus* possess a lesser number of these two amino acids compared to other mosquitoes. Among VgC1 and VgC2 of *Cx. quinquefasciatus* the reduction is more prominent in the truncated gene of VgC2.

In addition, the protein's 3D structure (Fig. 4) was also predicted by EsyPred 3D. One can clearly see the beta-sheet shell regions and the superhelical domains that are typical of lipovitellin. The predicted isoelectric point (pI) and molecular weight (MW) of Vg were 5. 64 and 239 694. 44 Da, respectively, through Compute pI/MW.

```
MIAKLLLLTFVGLCTAYQYSYEYEFPSSRPFNKTGFEFGAWEPNREYVYNVTTKTMTALPDLEDYWTGIVTHGYLVIRPK      80
DHNYVVAYIDRPMYAAFNEYLPRGYRTELSHFNLKWQPMPFSSKPFGIYYNKGAVKGFYVEKSVPNHEVNMLKGWVSQFQ     160
LDTQGAYVIKSEFNQFPENNTLTGVYKTMEPSVTGECETLYDVNPVPEYHFQSHKEWVPQPQWLEEDQHVFHVVKSRNFD     240
RCEQRMGFHFGFSGFSDFKPNTNQMGNIMTKSEVTQMYLTGNWYNYTIQSVSTVNKVVVSPSLVNSQKAMSTLQVNMTLN     320
EITPYNKYPEGPADDRQVFVDLVYSYNMAHDKKNFVRPANETDDSSSSSSDSSSDSSSSSDSSSSSEEEHENFKISPSEQ     400
YKKQVKEVEHRGNRNRRDLNAFKEKQYYESYKRDQYRLRKHNDTSSDSSSSDDSNSSSSSSSSDESDEHDFYSSSESDSH     480
SLSSSEEDFYQPIPESMKEAPQTPFLPYFTGYKGYSVQYAHNVDASRYAYKLAYEIAEELQEMSQVPKSNTLNKFTILARV     560
LRTMHYQDIYDVCQKLFVSQKEREEGSNHSESFAKKCDAWNTFRDALAQAGTPPAFKVIKELIEEKKLRGDEAASVIATL     640
PKTIRYPTETVMHEYFLLVTSNAVQHQEYLNTTALISFCDFLNRAQVNNRSAYNYYPVHSFGRLADADYKIVAHKVVPWF     720
AHQLREAVKAGDSVKVQVYIRCLGHLGHPEILNVFEPYLEGKIPVTHFQRLAFIVALDRLVENYPRLARSVLFKVYQNTG     800
DAHEVRCAAVYLLIRTKPPVYMLQRMAEQTHYDPSTYVRAAVKTALESASEADEFDDDYEFSQNAQAAIKHLNPRDFSLQ     880
YSGTYLRDFAFKELELSYRMYFSQIAADDHYVPSGFFFHLRKNMGGLKRFSTFYYLVSSMETFFDLLDKQYDSYNKHSEY     960
KSSDYYYKYYKQYPSLFKDYFSQYSKNHKYQNDYYEQFGNKNQEEFQKWSTTRIAKLLNIDPEEAEELEGQFMFQIFNGE    1040
RFFAFNNQTIEQFPSFVKKYFEDFEDGFAYNVTKFYQQNVVTMAFPLATGLPFTYSLKTPTLMKFEFEATATTHPSIYKT    1120
PTGYPEKEYDDFIHLPRWFNGSADVNMAYSRLVDAKVGFITPFDHQRYVAGYQKKFQGYLPFSFDFGFDFENNDFEVNVQ    1200
PLEPKKDALLFHMSSWPYTGYKDITDLRPMAEQPSVHILHDRAQTTKSFETSFGHELTGVALRFQAKYDKDFIDYAYLMK    1280
HIEQHDYWSALVYPFASETYHYHQLNLYYDAQRTSVKNVKFVLQHKQADYDQDFQTADVKHPKGRHGFSGYYNEFNYAQP    1360
FVYYAGSQRRQEQFMRNAGAGIRNSDVNVFDFGIVFEGKQQKAEFVFTTAYADSPVDEKERLLMFLSFSPYVSSSAFYEF    1440
IPFSGKQFQMCFSATNQYPNMPKLNFLNVLNFDKIGSMDWELSYGEKCQGGSHVSMKGKLIQSEPYRHFLRISEAGQRCK    1520
QQMDKGYFQLPACQNATRQAGYFDQYSFNFEYKDVSNYAKNLTYQFFDYARYFSFPYWSEDYFFQGKHNQFQIDFQLAPY    1600
FDYYNASFYGTDRSFAIQNYPIESEYARYFFSIHPDFDYYERMFNYAYRGNYHPSCVVSNKFVNTFDGKTYDYELGNCWH    1680
VVLHTVKPDYYFYAQDSHFMNSDYEYNWKNGFGEDEQITILARHGEDNQLFLKAILGQYKQNDYNIDIIPHGHELPMVYI    1760
NGKPQQIHEKYAVEMYTNDDGGDQPLIRVYALPGNELEISFRDDDIKIVFDGYRARFFADQSYFNNFVGLCGTNNGEGED    1840
DFITPDQCVMRKPEYFAASYALSGMNCSGPAQAYFTEYHQKAQEHCVKPQYYFGNVISEQEAGRQRYNYYYKDFDLSDSS    1920
SSESSSSSSSSESDESDDSNSSSSEEQKPNREHFFEKQQYTEKECPVKHQAQYVEQGDKICFTSRPLPACASQCKATEKV    2000
PKYVDVHCRDVTDSVAQLYKQQIRKGVNPDMSNKSVTKTVKFFLPKKCVHVY
```

Fig. 3 The deduced amino acid sequence of *AncuVg*

Amino acids are numbered from the translation initiation methionine on the right. The first 16 residues with an underline constitute the predicted signal peptide. Three polyserine regions are indicated by yellow highlighting, while RXXR motifs are shown by wavy lines. DGXR and GLCG motif are shown with a thick underline. The nine conserved cysteines are indicated with bold and underlined letters. The twenty putative glycosylation sites are shown in blue color while the amino acids shown in red are asparagine linked glycosylation sites. The phosphoryaltion sites, serines, threonines and tyrosines are marked with dotted lines.

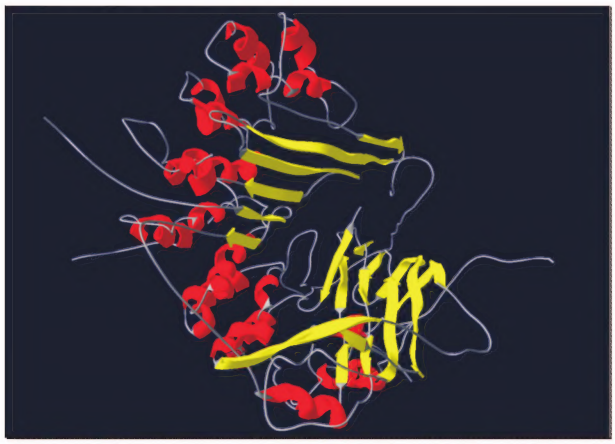

Fig. 4 Ribbon representation of the 3D model of the VG protein

The helices are depicted in red, beta strand in yellow and coil in gray.

3.4 Synonymous codon usage in autogenous and anautogenous mosquito vitellogenins

The extent of codon bias in mosquito vitellogenin genes based on the relative synonymous codon usage (RSCU) values was determined by the effective number of codons (ENC) (data not shown). Vitellogenin genes of mosquitoes have a codon bias that ranged from ENC 32.3 (*A. albimanus*) to 50.7 (*Tx. amboinensis*); the larger the extent of codon preference in a gene, the smaller the corresponding ENC value. Pattern of synonymous codon usage in the autogenous and anautogenous mosquito vitellogenin genes was examined (data not shown). A high synonymous codon usage bias has been revealed from Vg genes of all anautogenous mosquitoes, particularly using one or two synonymous codons over others. Amino acids like phenylalanine, tyrosine, glutamine, asparagine, lysine and cysteine are almost solely encoded by one synonymous codon. Mostly the preferential codons were found to be conserved in all the anautogenous mosquito vitellogenin genes during analysis for interspecific codon choice, except for two amino acids in *Ae. aegypti* where GAG is preferred over GAA for glutamic acid and GAT is preferred over GAC for aspartic acid.

The autogenous mosquitoes (*Oc. atropalpus* and *Tx. amboinesis*) showed the lowest synonymous codon usage bias in all vitellogenin genes analyzed. It was interesting to note that codon usage preference

for glutamic acid and aspartic acid was similar in both autogenous mosquitoes and *Ae. aegypti*. The average ENC for the seven anautogenous species was 38.0, and for the two autogenous species was 47.4. It was found that there was a positive correlation between GC content and the degree of synonymous codon usage bias measured by ENC (data not shown). The highest synonymous codon usage bias of *A. albimanus* VgC with an ENC of 32.3 indicates the biased GC_{3s} (83.5%).

3.5 The promoter region

In the present study, the upstream region of the *AncuVg* gene contains four regions for E74, three for BRC Z2, two for BRC Z3 and three for BRC Z4 response elements. In addition to these, the 5′ upstream region of the gene also contains four EcRE like sequences that match 9/13, 12/14, 11/13 and 12/14 bp (Fig. 5). The upstream region of the gene also contains binding sites for GATA, C/EBP and HNF-3 transcription factors. The transcription factor binding sites were also confirmed by CONSITE software in addition to the manual searching. However, to confirm the roles of the regulatory factors in Vg gene transcription, further analysis is required.

Multiple sequence alignments of regulatory sequence of mosquito vitellogenins were performed by the CLUSTAL W program as shown in Fig. 6. Comparisons of the regulatory sequences of vitellogenin genes showed that the pattern of conservation is different between *Anopheles* species and other mosquito species. The GATA rich region was found conserved in *Anopheles* at 187 nucleotide position whereas a similarly less conserved region was found in *Aedes* at a downstream location. Similarly, an AT rich region was found almost conserved in *Anopheles* species at 218 position, however, no such conservation was found in other mosquito species examined.

The TATA box was found almost conserved in all the mosquitoes except for vitellogenin 2a and 2b of *Cx. tarsalis*. The transcription start site was found positionally conserved in seven species of anautogeneous mosquitoes as well as in two species of autogenous mosquitoes. A very small thymidine block, downstream of transcription start site was found almost conserved in ananutogenous mosquitoes however; the same block was not found conserved in autogenous mosquitoes. Similarly, one more thymidine block was found just adjacent to the previous one in all mosquitoes examined except *Anopheles* mosquitoes. The translation start site was found fully conserved in all ananutogenous or autogenous mosquitoes.

3.6 Phylogenetic analysis

Phylogenetic analysis using the 5′ upstream region of Vg genes was examined for evolutionary patterns among mosquitoes (Fig. 7). All analyses gave identical tree morphology conforming to three major clades: *Anopheles*, *Aedes/Ochlerotatus*, and *Culex*, with the exception of the *Culex* Vg2 group (which represents a duplication event unique to the

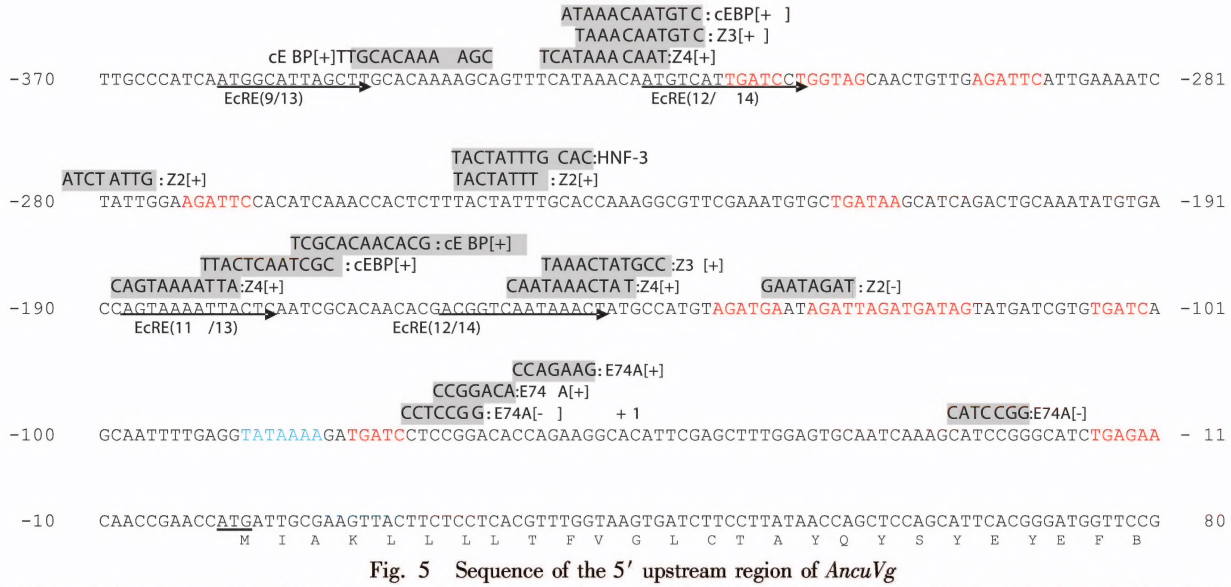

Fig. 5 Sequence of the 5′ upstream region of *AncuVg*

Bold letters indicate the putative binding sites for ecdysone response element EcRE, E74A, cEBP, HNF/3 and BRC Z2 (Z1) Z3 (Z3) and Z4 (Z4) isoforms. The GATA binding sites are shown in red while the TATA box is shown in blue. Direction is indicated by the arrows. The transcription initiation site is designated as +1. The ATG initiation codon is indicated with a thick underline.

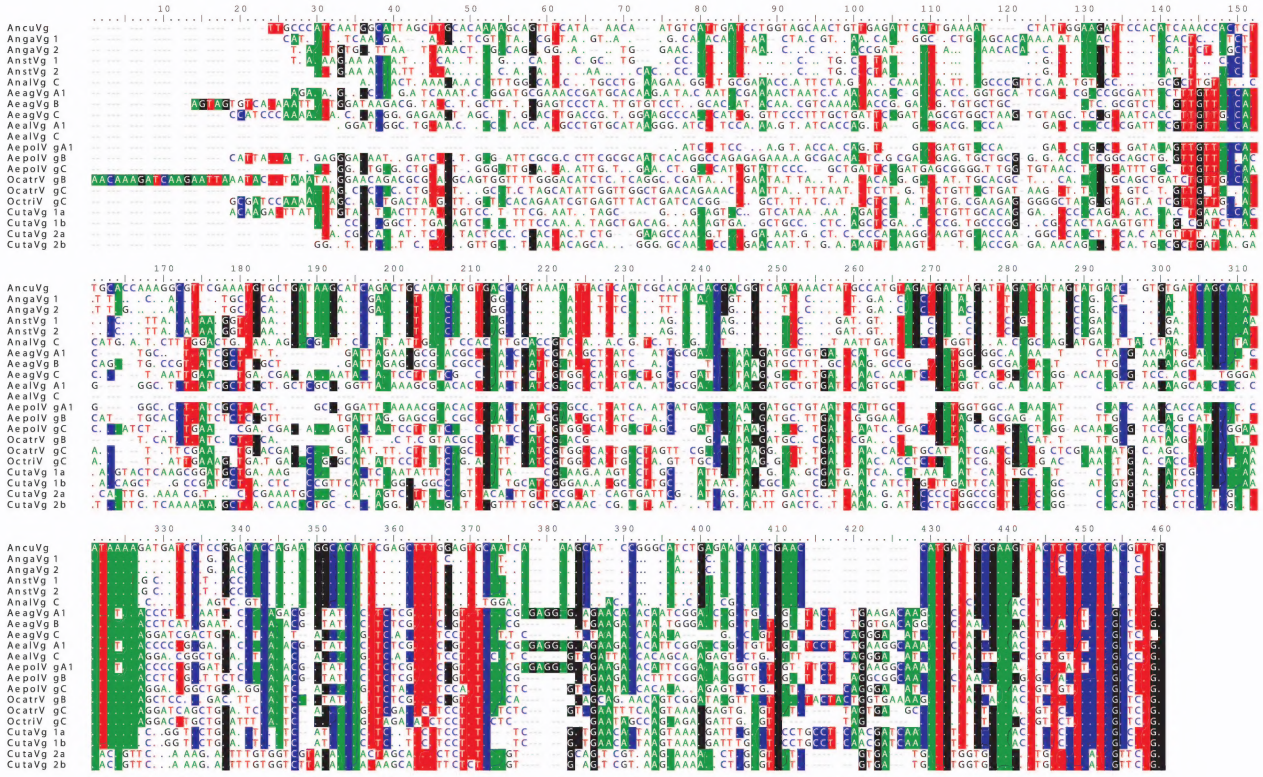

Fig. 6 Alignment of the putative *cis*-acting regulatory elements of mosquito vitellogenin-encoding gene promoters

Twenty-one sequences corresponding to vitellogenin-encoding genes were aligned to evaluate sequence identities and similarities. Conserved residues are highlighted in their respective color alignments, (……) indicate similar sequences, (------) indicates missing sequences.

genus *Culex*). The *Aedes/Ochlerotatus* clade is more closely related to the *Culex* Vg1 clade than the *Anopheles* clade, which is in agreement with the agreed relationships of these genera. Amongst the *Anopheles* group, *A. culicifacies* is found to be placed nearer to *A. gambiae* than to *A. stephensi*.

4 DISCUSSION

Vitellogenin is the major yolk protein precursor which is synthesized extensively after blood feeding in the fat body of the female mosquitoes. In the present study, the Vg gene of *A. culicifacies* shows strong nucleotide similarity to *A. gambiae* and *A. stephensi* as compared with other mosquito Vg genes (Table 2). The *AncuVg* gene also comprises 3 exons and 2 introns as found in other mosquitoes. The size of the 1st exon (31 bp followed by ATG) was found to be constant in the Vg genes of all mosquitoes (Table 3). However, the variations in size and position of other exons and introns indicate the divergence of Vg genes among mosquitoes. The intron boundaries of *AncuVg* has conserved splice sites. The presence of repetitive DNA elements may be the reason for the variation in the size and position of introns as observed among vertebrates (Gerber-Huber *et al.*, 1987).

The certain characteristic features of an insect vitellogenin protein sequence (Chen *et al.*, 1997; Sappington and Raikhel, 1998; Sappington *et al.*, 2002) were also found in *AncuVg* amino acid sequence. These include four consensus cleavage sites (RXXR motifs in different parts of the primary product), and highly conserved motif (GL/ICG) followed by nine cysteine residues at conserved locations near the C-terminus (Chen *et al.*, 1997; Comas *et al.*, 2000; Lee *et al.*, 2000a, 2000b; Tufail *et al.*, 2001). The GL/ICG motif and cysteine residues play an important role in the oligomerization of vertebrate vitellogenins (Mayadas and Wagner, 1992; Mouchel *et al.*, 1996). In addition, a DGXR motif which occurs in almost all insect vitellogenins studied to date (Tufail *et al.*, 2000) was located 17 – 19 residues upstream of the GL/ICG motif in *AncuVg* gene. The role of DG residues of DGXR motif along with GL/ICG motif and cysteine residues are hypothesized to form a structure which may be necessary for proper functioning of vitellogenins during embryogenesis (Tufail *et al.*, 2001).

The AncuVg protein sequence also contains signal peptides of 16 amino acids at the 5′ terminus, an N_Vitellogenin region and a VWF type D domain as also found conserved in other insect Vgs (Romans *et al.*, 1995; Lee *et al.*, 2000a; Tufail and Takeda,

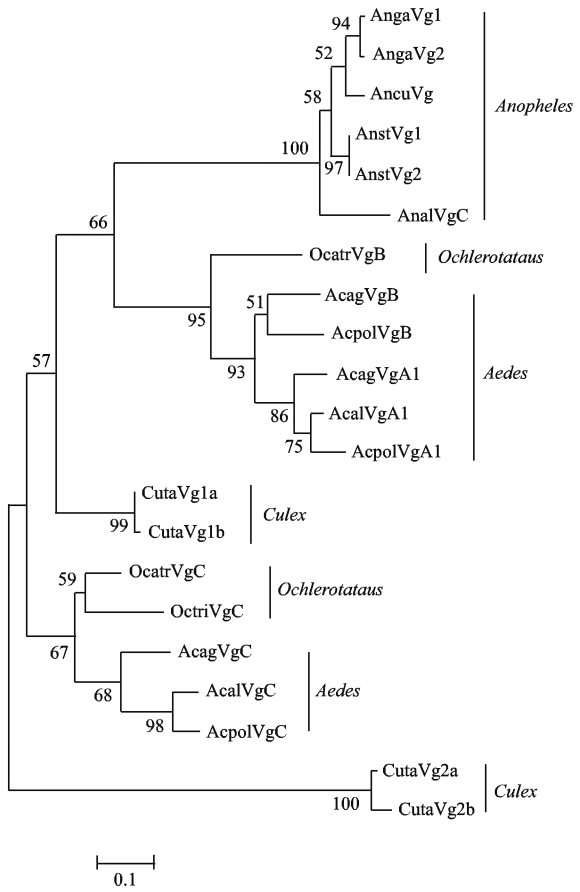

Fig. 7　Phylogenetic analysis of mosquitoes based on regulatory sequences of Vg genes

The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 3.14540346 is shown. The percentages of replicate trees in which the associated taxa are clustered together in the bootstrap test (1 000 replicates) are shown next to the branches. Evolutionary analyses were conducted in MEGA 5.

2008). The existence of three polyserine tracts in *A. culicifacies* Vg, two in the amino-terminus and one in the carboxy-terminus was similar to those observed in several other Vgs (Romans *et al.*, 1995; Sappington and Raikhel, 1998).

The deduced molecular weight of the Vg protein was observed as 238.0 kDa having four putative cleavage sites at the amino acids R_{103}GYR, R_{411}GNR, R_{559}VLR and R_{786}LAR (Fig. 2). The cleavage sites are immediately preceded by a motif, (R/K) X (R/K) R or RXXR (reviewed by Sappington and Raikhel, 1998). Accordingly the putative protein could be cleaved into nine subunits of approximately 228, 192, 175, 89, 78, 63, 52, 46 and 43 kDa. However, the exact molecular weight of the native Vg protein of *AncuVg* is yet to be validated. Most insect Vg mRNAs are about 6 – 7 kb long encoding about 1 800 amino acids in a single open reading frame (Tufail and Takeda, 2008). The weights of Vg proteins are >200 kDa in most insect species and the primary translation product of many insect Vg genes is generally cleaved into larger (>150 kDa) and smaller (<65 kDa) subunits (Chen *et al.*, 1997; Hagedorn *et al.*, 1998; Sappington and Raikhel, 1998; Raikhel *et al.*, 2005).

The post translational modifications, *i.e.*, putative glycosylation sites and predicted phosphorylated residues, were similar as that observed earlier in all insects (Raikhel and Dhadialla, 1992; Hagedorn *et al.*, 1998; Giorgi *et al.*, 1999; Tufail *et al.*, 2005). Glycosylation is an important step in the subsequent secretion of Vg by the fat body (Wyatt *et al.*, 1984; Wojchowski *et al.*, 1986; Dhadialla and Raikhel, 1990; Don-Wheeler and Engelmann, 1997). However, high level of phosphorylation especially at serine residues may be primarily due to the existence of polyserine tracts at both the termini which are the most flexible domains shielded from proteolysis by phosphorylation (Havukainen *et al.*, 2012).

Specific amino acid usage has been revealed by amino acid composition analysis of mosquito vitellogenin proteins. All mosquito vitellogenins possess higher content of serine residues due to the presence of three polyserine tracts. Two aromatic amino acids, *i.e.*, tyrosine and phenylalanine were also found in higher content in all mosquito vitellogenins except in the genera *Culex* for some unknown reasons. A significantly high synonymous codon usage bias has been shown by mosquito vitellogenin genes, by predominantly using one or two optimal synonymous codons over others. Generally, highly expressed genes have much more selective constraints on synonymous codon choices for translational competency. The results acquired here also revealed the similar pattern as observed from other highly expressed genes (Ikemura 1985; Kurland, 1991). The chief factor responsible for affecting the synonymous codon usage bias in the mosquito vitellogenin genes could be the amount of GC content at the 3rd codon position as in the case of *D. melanogaster* (Shields *et al.*, 1988; Powell and Moriyama, 1997). However, lower synonymous codon usage bias has been observed in vitellogenin genes of two autogenous mosquito species, *Oc. atropalpus* and *Tx. amboinensis* due to low values of GC content at the third codon position.

The analysis of the upstream region of the *AncuVg* gene revealed the presence of response elements for the ecdysone early genes *E74* and *BRC* (Z2, Z3 and Z4 isoforms) that are regulated by the ecdysone/EcR/USP complex. The E74 A and B isoforms bind to the same consensus sequence (Sun *et al.*, 2005); however, four isoforms (Z1 – Z4) encoded by the BRC gene has their own binding response elements (Bayer *et al.*, 1996). Similar binding sites for E74, E75 and BRC have also been reported in the regulatory region of *Ae. aegypti* gene (Kokoza *et al.*, 2001; Raikhel *et al.*,

2002). In *Ae. aegypti* E74 B (Kokoza *et al.*, 2001; Sun *et al.*, 2005) and BRC Z2 isoforms were found to be responsible for enhancing *AaVg* gene transcription while Z1 and Z4 isoforms of BRC gene were found to suppress *AaVg* gene transcription (Chen *et al.*, 2004; Zhu *et al.*, 2007). It has been confirmed that ecdysone/EcR/USP directly transactivates *AaVg* transcription in spite of the difference of actual binding sites from that of the consensus EcRE sequences (Kokoza *et al.*, 2001).

The presence of response elements for GATA (GATA transcription factor), C/EBP (CAAT-binding protein) and HNF3/fkh (hepatocyte nuclear factor 3/forkhead transcription factor) in the regulatory region of *AncuVg* could be involved in the correct tissue and stage specific expression as in *AaVg* (Kokoza *et al.*, 2001). GATA factor has also been shown to be involved in the amino acid/target of Rapamyin (TOR) pathway that regulates the nutrient dependent Vg transactivation (Hansen *et al.*, 2004; Park *et al.*, 2006).

Comparison of the 5′ upstream region of vitellogenin of all mosquitoes including the present study shows significant degree of conservation particularly in certain regions like TATA box, transcription start site and the translation start site. Unlike coding region, control region of vitellogenin gene from different mosquitoes shows high level of divergence. In general, the promoter region of mosquito vitellogenins was found to be A + T rich. High A + T content of this region might seem to be a phylogenetic characteristic of insecta, and also whether this high A + T content point towards any directional mutation pressure or not, that needs to be explored further. Inspite of sequence divergence and high A + T content of the control region, the phylogenetic tree reflects the same picture reported by earlier workers based on coding sequence of the gene.

The evolutionary patterns of insect vitellogenins based on phylogenetic trees have already been reviewed (Tufail and Takeda, 2008). Phylogenetic analysis using the 5′ upstream region in the present study confirmed the three major clades based on coding sequences: *Aedes/Ochlerotatus*, *Culex* and *Anopheles* (with the exception of the *Culex* Vg2 group) and also as expected *Aedes/Ochlerotatus* clade was found more closely related to *Culex* Vg1 clade than the *Anopheles* clade (Chen *et al.*, 2010). The present analysis also demonstrated that *A. culicifacies* is genetically closer to the *A. gambiae* and both the species would have been diverged at the same point of time from the common ancestor.

ACKNOWLEDGEMENTS Monika Miglani was awarded SRF by Haryana State Council for Science and Technology, Haryana.

References

Ahmed AM, Maingon R, Romans P, Hurd H, 2001. Effects of malaria infection on vitellogenesis in *Anopheles gambiae* during two gonotrophic cycles. *Insect Mol. Biol.*, 10: 347 – 356.

Bayer CA, Holley B, Fristrom JW, 1996. A switch in broad-complex zinc-finger isoform expression is regulated post transcriptionally during the metamorphosis of *Drosophila* imaginal discs. *Dev. Biol.*, 177: 1 – 14.

Byrne BM, Gruber M, Ab G, 1989. The evolution of egg yolk proteins. *Prog. Biophys. Mol. Biol.*, 53: 33 – 69.

Chen JS, Sappington TW, Raikhel AS, 1997. Extensive sequence conservation among insect, nematode, and vertebrate vitellogenins reveals ancient common ancestry. *J. Mol. Evol.*, 44: 440 – 451.

Chen L, Zhu J, Sun G, Raikhel AS, 2004. The early gene *Broad* is involved in the ecdysteroid hierarchy governing vitellogenesis of the mosquito *Aedes aegypti*. *J. Mol. Endocrinol.*, 33: 743 – 761.

Chen XG, Marinotti O, Whitman L, Jasinskiene N, James AA, Romans P, 2007. The *Anopheles gambiae* vitellogenin gene (*VGT2*) promoter directs persistent accumulation of a reporter gene product in transgenic *Anopheles stephensi* following multiple blood meals. *Am. J. Trop. Med. Hyg.*, 76(6): 1118 – 1124.

Chen S, Armistead JS, Provost-Javier KN, Sakamoto JM, Rasgon JL, 2010. Duplication, concerted evolution and purifying selection drive the evolution of mosquito vitellogenin genes. *BMC Evol. Biol.*, 10: 142.

Comas D, Piulachs MD, Belle' s X, 2000. Vitellogenin of *Blattella germanica* (L.) (Dictyoptera, Blattellidae): nucleotide sequence of the cDNA and analysis of the protein primary structure. *Arch. Insect Biochem. Physiol.*, 45: 1 – 11.

Dhadialla TS, Raikhel AS, 1990. Biosynthesis of mosquito vitellogenin. *J. Biol. Chem.*, 265: 9924 – 9933.

Don-Wheeler G, Engelmann F, 1997. The biosynthesis and processing of vitellogenin in the fat bodies of female and male of the cockroach *Leucophaea maderae*. *Insect Biochem. Mol. Biol.*, 27: 901 – 918.

Felsenstein J, 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evol.*, 39: 783 – 791.

Gakhar SK, Singh S, Shandilya H, 1997. Changes in soluble proteins during the development of malaria vector *An. stephensi* (Diptera: Insecta). *Proc. Natl. Sci. Acad. USA*, B63: 289 – 298.

Gakhar SK, Jhamb A, Gulia M, Dixit R, 2001. Anti-mosquito ovary antibodies reduce the fecundity of *Anopheles stephensi* (Diptera: Insecta). *Jpn. J. Infect. Dis.*, 54: 181 – 183.

Gerber-Huber SG, Nardelli D, Haefliger JA, Cooper DN, Givel F, Germond JE, Engel J, Green NM, Wahli W, 1987. Precursor-product relationship between viteliogenin and the yolk proteins as derived from the complete sequence of a *Xenopus* viteliogenin gene. *Nucleic Acids Res.*, 15: 12.

Giorgi F, Bradley JT, Nordin JH, 1999. Differential vitellin polypeptide processing in insect embryos. *Micron*, 30: 579 – 596.

Green CA, Miles SJ, 1980. Chromosomal evidence for sibling species of the malaria vector *Anopheles* (*Cellia*) *culicifacies* Giles. *J. Trop. Med. Hyg.*, 83: 75 – 78.

Hagedorn HH, Maddison DR, Tu Z, 1998. The evolution of vitellogenins, cyclorrhaphan yolk proteins and related molecules. *Adv. Insect Physiol.*, 27: 335 – 384.

Hansen IA, Attardo GM, Park JH, Peng Q, Raikhel AS, 2004. Target of rapamycin-mediated amino acid signaling in mosquito anautogeny. *Proc. Natl. Sci. Acad. USA*, 101: 10626 – 10631.

Havukainen H, Underhaug J, Wolschin F, Amdam G, Halskau Ø, 2012. A vitellogenin polyserine cleavage site: highly disordered conformation protected from proteolysis by phosphorylation. *J. Exp. Biol.*, 215: 1837 – 1846.

Ito J, Ghosh A, Moreira LA, Wimmer EA, Jacobs-Lorena M, 2002. Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature*, 417: 452 – 455.

Ikemura T, 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.*, 2: 13 – 34.

Isoe J, Hagedorn HH, 2007. Mosquito vitellogenin genes: comparative

sequence analysis, gene duplication, and the role of rare synonymous codon usage in regulating expression. *J. Insect Sci.*, 7: 1 - 49.

Kar I, Subbarao SK, Eapen A, Ravindran J, Satyanarayana TS, Raghavendra K, Nanda N, Sharma VP, 1999. Evidence for a new malaria vector species, species E, within the *Anopheles culicifacies* complex (Diptera: Culicidae). *J. Med. Entomol.*, 36: 595 - 600.

Kokoza VA, Martin D, Mienaltowski MJ, Ahmed A, Morton CM, Raikhel AS, 2001. Transcriptional regulation of the mosquito vitellogenin gene via a blood meal-triggered cascade. *Gene*, 274: 47 - 65.

Kurland CG, 1991. Codon bias and gene expression. *FEBS Letters*, 285: 165 - 169.

Lambert C, Léonard N, De Bolle X, Depiereux E, 2002. ESyPred3D: Prediction of proteins 3D structures. *Bioinformatics*, 18 (9): 1250 - 1256.

Lee JM, Hatakeyama M, Oishi K, 2000a. A simple and rapid method for cloning insect vitellogenin cDNAs. *Insect Biochem. Mol. Biol.*, 30: 189 - 194.

Lee JM, Nishimori Y, Hatakeyama M, Bae TW, Oishi K, 2000b. Vitellogenin of the cicada *Graptopsaltria nigrofuscata* (Homoptera): analysis of its primary structure. *Insect Biochem. Mol. Biol.*, 30: 1 - 7.

Li J, Huang J, Cai W, Zhao Z, Peng W, Wu J, 2010. The vitellogenin of the bumblebee, *Bombus hypocrita*: studies on structural analysis of the cDNA and expression of the mRNA. *J. Comp. Physiol. B*, 180(2): 161 - 170.

Mayadas TN, Wagner DD, 1992. Vicinal cysteines in the prosequence play a role in von Willebrand factor multimer assembly. *Proc. Natl. Sci. Acad. USA*, 89: 3531 - 3535.

Moreira LA, Ito J, Ghosh A, Devenport M, Zieler H, Abraham EG, Crisanti A, Nolan T, Catteruccia F, Jacobs-Lorena M, 2002. Bee venom phospholipase inhibits malaria parasite development in transgenic mosquitoes. *J. Biol. Chem.*, 277: 40839 - 40843.

Mouchel N, Trichet V, Betz A, Le Pennec JP, Wolff J, 1996. Characterization of vitellogenin from rainbow trout (*Oncorhynchus mykiss*). *Gene*, 174: 59 - 64.

Nielsen H, Engelbrecht J, Brunak S, von Heijne G, 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.*, 10: 1 - 6.

Nirmala X, Marinotti O, James AA, 2005. The accumulation of specific mRNAs following multiple blood meals in *Anopheles gambiae*. *Insect Mol. Biol.*, 14: 95 - 103.

Nirmala X, Marinotti O, Sandoval JM, Phin S, Gakhar S, Jasinskiene N, James AA, 2006. Functional characterization of the promoter of the vitellogenin gene, *AsVg*1, of the malaria vector, *Anopheles stephensi*. *Insect Biochem. Mol. Biol.*, 36(9): 694 - 700.

Nose Y, Lee JM, Ueno T, Hatakeyama M, Oishi K, 1997. Cloning of cDNA for vitellogenin of the parasitoid wasp *Pimpla nipponica* (Hymenoptera: Apocrita: Ichneumonidae): vitellogenin primary structure and evolutionary considerations. *Insect Biochem. Mol. Biol.*, 27: 1047 - 1056.

Orkin SH, 1992. GATA-binding transcription factors in hematopoietic cells. *Blood*, 80: 575 - 581.

Park JH, Attardo GM, Hansen IA, Raikhel AS, 2006. GATA factor translation is the final downstream step in the amino acid/target-of-rapamycin-mediated vitellogenin gene expression in the anautogenous mosquito *Aedes aegypti*. *J. Biol. Chem.*, 281: 11167 - 11176.

Powell JR, Moriyama EN, 1997. Evolution of codon usage bias in *Drosophila*. *Proc. Natl. Sci. Acad. USA*, 94: 7784 - 7790.

Puigbo P, Bravo IG, Garcia-Vallve S, 2008. CAIcal: a combined set of tools to assess codon usage adaptation. *Biol. Dir.*, 3: 38.

Raikhel AS, Dhadialla TS, 1992. Accumulation of yolk proteins in insect oocytes. *Annu. Rev. Entomol.*, 37: 217 - 251.

Raikhel AS, Kokoza VA, Zhu J, Martin D, Wang SF, Li C, 2002. Molecular biology of mosquito vitellogenesis: from basic studies to genetic engineering of antipathogen immunity. *Insect Biochem. Mol. Biol.*, 32: 1275 - 1286.

Raikhel AS, Brown MR, Belles X, 2005. Hormonal control of reproductive process. In: Gilbert LI, Iatrou K, Gill SS eds. Comprehensive Molecular Insect Science. Elsevier Limited, Oxford. 433 - 491.

Richard JE, Lin K, Tan T, 2000. A functional significance for codon third bases. *Gene*, 245: 291 - 298.

Riehle MA, Srinivasan P, Moreira CK, Jacobs-Lorena M, 2003. Towards genetic manipulation of wild mosquito populations to combat malaria: advances and challenges. *J. Exp. Biol.*, 206: 3809 - 3816.

Romans P, Tu Z, Ke Z, Hagedorn HH, 1995. Analysis of a vitellogenin gene of the mosquito, *Aedes aegypti* and comparisons to vitellogenins from other organisms. *Insect Biochem. Mol. Biol.*, 25: 939 - 958.

Saitou N, Nei M, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 4: 406 - 425.

Sambrook J, Fritsch EF, Maniatis T, 1989. Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sandelin A, Wasserman WW, Lenhard B, 2004. ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acid Res.*, 1: 32.

Sappington TW, Raikhel AS, 1995. Receptor-mediated endocytosis of yolk proteins by insect oocytes. In: Takahashi S ed. Recent Advances in Insect Biochemistry and Molecular Biology. Nagoya University Press, Nagoya, Japan. 235 - 257.

Sappington TW, Raikhel AS, 1998. Molecular characteristics of insect vitellogenins and vitellogenin receptors. *Insect Biochem. Mol. Biol.*, 28: 277 - 300.

Sappington TW, Oishi K, Raikhel AS, 2002. Structural characteristics of insect vitellogenins. In: Adiyodi KG, Adiyodi RG eds. Reproductive Biology of Invertebrates, Vol. 12, Part A. Science Publishers, Inc., Enfield, New Hampshire. 69 - 101.

Sharp PM, Tuohy TMF, Mosurski KR, 1986. Codon usage in yeast cluster-analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.*, 14: 5125 - 5143.

Shields DC, Sharp PM, Higgins DG, Wright F, 1988. "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Mol. Biol. Evol.*, 5: 704 - 716.

Subbarao SK, Vasantha K, Adak T, Sharma VP, 1983. *Anopheles culicifacies* complex: evidence for a new sibling species, species C. *Ann. Entomol. Soc. Am.*, 76: 985 - 986.

Suguna SG, Tewari SC, Mani TR, Hiryan J, Reuben R, 1989. A cytogenetic description of a new species of the *Anopheles culicifacies* species complex. *Genetica*, 78: 225 - 230.

Sun G, Zhu J, Chen L, Raikhel AS, 2005. Synergistic action of E74B and ecdysteroid receptor in activating a 20-hydroxyecdysone effector gene. *Proc. Natl. Acad. Sci. USA*, 102: 15506 - 15511.

Tamura K, Nei M, Kumar S, 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Sci. Acad. USA*, 101: 11030 - 11035.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S, 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*, 28(10): 2731 - 2739.

Thompson JD, Higgins DG, Gibson TJ, 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 11; 22(22): 4673 - 4680.

Tufail M, Lee JM, Hatakeyama M, Oishi K, Takeda M, 2000. Cloning of vitellogenin cDNA of the American cockroach, *Periplaneta americana* (Dictyoptera), and its structural and expression analyses. *Arch. Insect Biochem. Physiol.*, 45: 37 - 46.

Tufail M, Hatakeyama M, Takeda M, 2001. Molecular evidence for two vitellogenin genes and processing of vitellogenins in the American cockroach, *Periplaneta americana*. *Arch. Insect Biochem. Physiol.*, 48: 72 - 80.

Tufail M, Raikhel AS, Takeda M, 2005. Biosynthesis and processing of insect vitellogenins. In: Raikhel AS, Sappington TW eds. Progress in Vitellogenesis. Reproductive Biology of Invertebrates, Vol. XII. Part B. Science Publishers, Inc., Enfield (NH), USA/Plymouth,

UK. 1 – 32.
Tufail M, Bembenek J, Elgendy AM, Takeda M, 2007. Evidence for two vitellogenin-related genes in *Leucophaea maderae*: the protein primary structure and its processing. *Arch. Insect Biochem. Physiol.*, 66: 190 – 203.
Tufail M, Takeda M, 2008. Molecular characteristics of insect vitellogenins. *J. Insect Physiol.*, 54: 1447 – 1458.
Urness LD, Thummel CS, 1995. Molecular analysis of a steroid-induced regulatory hierarchy: the *Drosophila* E74A protein directly regulates L71-6 transcription. *EMBO J.*, 14: 6239 – 6246.
Vasantha K, Subbarao SK, Sharma VP, 1991. *Anopheles culicifacies* complex: population cytogenetic evidence for species D (Diptera: Culicidae). *Ann. Entomol. Soc. Am.*, 84: 531 – 536.
von Kalm L, Crossgrove K, Von Seggern D, Guild GM, Beckendorf SK, 1994. The Broad-Complex directly controls a tissue-specific response to the steroid hormone ecdysone at the onset of *Drosophila metamorphosis*. *EMBO J.*, 13: 3505 – 3516.
World Health Organisation (WHO), 2011. World Malaria Report.
Wojchowski DM, Parsons P, Nordin JH, Kunkel JG, 1986. Processing of provitellogenin in insect fat body: a role for high mannose oligosaccharides. *Dev. Biol.*, 116: 422 – 430.
Wright F, 1990. The effective number of codons used in a gene. *Gene*, 87: 23 – 29.
Wyatt GR, Locke J, Bradfield JY, 1984. The vitellogenin genes for Locusta migratoria and other insects. In: Engles W ed. Advances in Invertebrate Reproduction, Vol. 3. Elsevier, Amsterdam. 73 – 80.
Zhu J, Chen L, Raikhel AS, 2007. Distinct roles of Broad isoforms in regulation of the 20-hydroxyecdysone effector gene, vitellogenin, in the mosquito *Aedes aegypti*. *Mol. Cell. Endocrinol.*, 267: 97 – 105.

印度疟疾媒介库态按蚊卵黄蛋白原基因的克隆与生物信息学分析

Monika MIGLANI, Surendra Kumar GAKHAR *

(Centre for Biotechnology, Maharshi Dayanand University, Rohtak, Haryana 124001, India)

摘要: 卵黄蛋白原(vitellogenin, Vg)是主要的卵黄蛋白前体, 在雌虫血餐之后在脂肪体内大量合成。卵黄蛋白原的调节元件已经被用于驱动蚊子(与寄生虫发生最大相互作用的场所)中抗寄生基因的组织特异性表达。不过, 迄今为止, 对在印度引起60% ~70%疟疾发生的库态按蚊 *Anopheles culicifacies* 中的内源启动子尚未进行过分析。本研究通过 PCR 扩增了包括 5′端上游调节区在内的库态按蚊 *A. culicifacies* 卵黄蛋白原基因, 并命名为 *AncuVg* (GenBank 登录号为 JN113091)。它含有一个大约 6.2 kb 的开放阅读框, 编码 2 052 个氨基酸, 具有一个 16 个氨基酸残基的推断的信号肽。也含有一个 N_Vitellogenin 区和一个 VWF 型 D 区, 这两个区在其他昆虫卵黄蛋白原中也保守。估计多肽分子量为 238.0 kDa, 含有 4 个共有的(RXXR/S)切割位点, C 端附近有一个 GL/ICG 基序, 其后是 9 个半胱氨酸残基和 1 个位于 GL/ICCG 基序上游第 18 个氨基酸残基处的 DGXR 基序。在推断的氨基酸序列上发现 3 个聚丝氨酸区, 其中 2 个位于氨基端, 1 个位于羧基端。根据同义密码子相对使用概率值, 通过有效密码子数, 测定了蚊子卵黄蛋白原基因密码子的偏倚性程度。也预测了库态按蚊 *A. culicifacies* Vg 的三维结构。分析了 *AncuVg* 基因, 以理解 Vg 基因的转录调节。对 Vg 基因 5′端上游区进行的系统发育分析表明, 它们聚类于蚊子的 3 大分枝。也用各种生物信息学工具分析分析了 Vg 的同源性和特征。

关键词: 库态按蚊; 卵黄蛋白原; 蚊子; 系统发育分析; 聚丝氨酸; 转录因子

中图分类号: Q966 **文献标志码**: A **文章编号**: 0454-6296(2013)09-1063-12

(责任编辑: 赵利辉)